DISCUSSION ON EFFICIENT DESIGNS

Anastasios A. Tsiatis
North Carolina State University

http://www.stat.ncsu.edu/∼tsiatis/

Outline

- Surrogate markers
- Group-sequential methods
- Auxiliary covariates
- Powering study for large effects

Gaining Efficiency through Surrogate Markers (Tom Fleming)

- Randomized clinical trials are used to evaluate new treatments in patients with chronic disease
- Primary endpoint is often clinical progression
- Such studies may
  - take too long
  - too many patients
  - problem with compliance and drop-outs
- An acceptable surrogate marker
  - responds rapidly to treatment
  - response implies a benefit regarding clinical outcome

Advantages of Using a Surrogate Marker Trial

- completed in shorter time
- fewer patients

Disadvantages

- Yet use of an invalid/incomplete surrogate marker can severely protract identification and development of effective therapies
- Important to assess potential surrogate markers for empirical validity, understand its limitations, and assess the uncertainties and implications of use of such a marker to predict a new drug's clinical efficacy

Assessing Surrogate Markers in HIV Clinical Trials

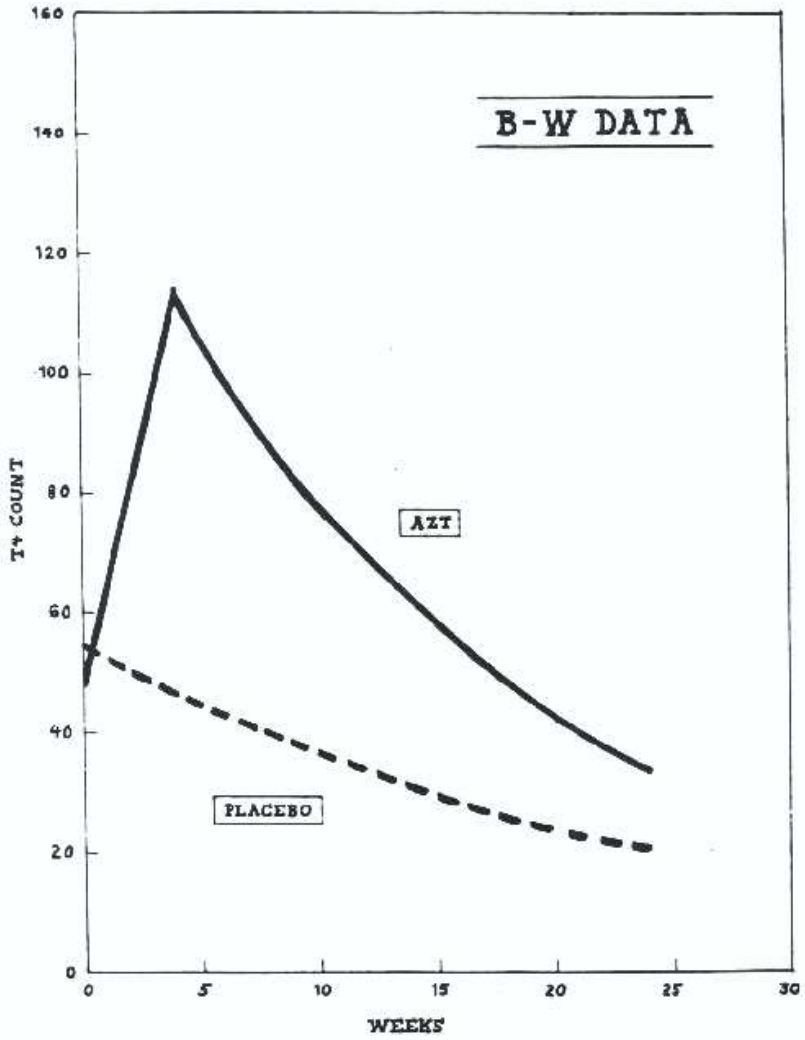
- For patients infected with HIV disease potential surrogate markers are
  - CD4 counts (a measure of the destruction in the immune system)
  - HIV_1 RNA (measures levels of circulating virus)

Prentice Definition of surrogate marker

1. Treatment should effect the surrogate marker; that is the values of the marker should be greater for "treatment group" than for "placebo group"
2. The marker should be prognostic; that is, the risk of progression should decrease as the value of the marker increases
3. The effect of treatment should manifest itself through the marker. i.e. The risk of progression given a specific marker response should be independent of treatment

CD4 Counts as a Potential Surrogate Marker

- Let us consider the role of CD4 counts as a potential surrogate marker
- BW 02
- time to death among patients with AIDS
- 281 patients with advanced HIV disease; 137 placebo, 144 ZDV
- median duration of 120 days
- study was stopped after seven months


B-W DATA
T+ COUNT
WEEKS
AZT
PLACEBO
160
140
120
100
80
60
40
20
0
0
5
10
15
20
25
30

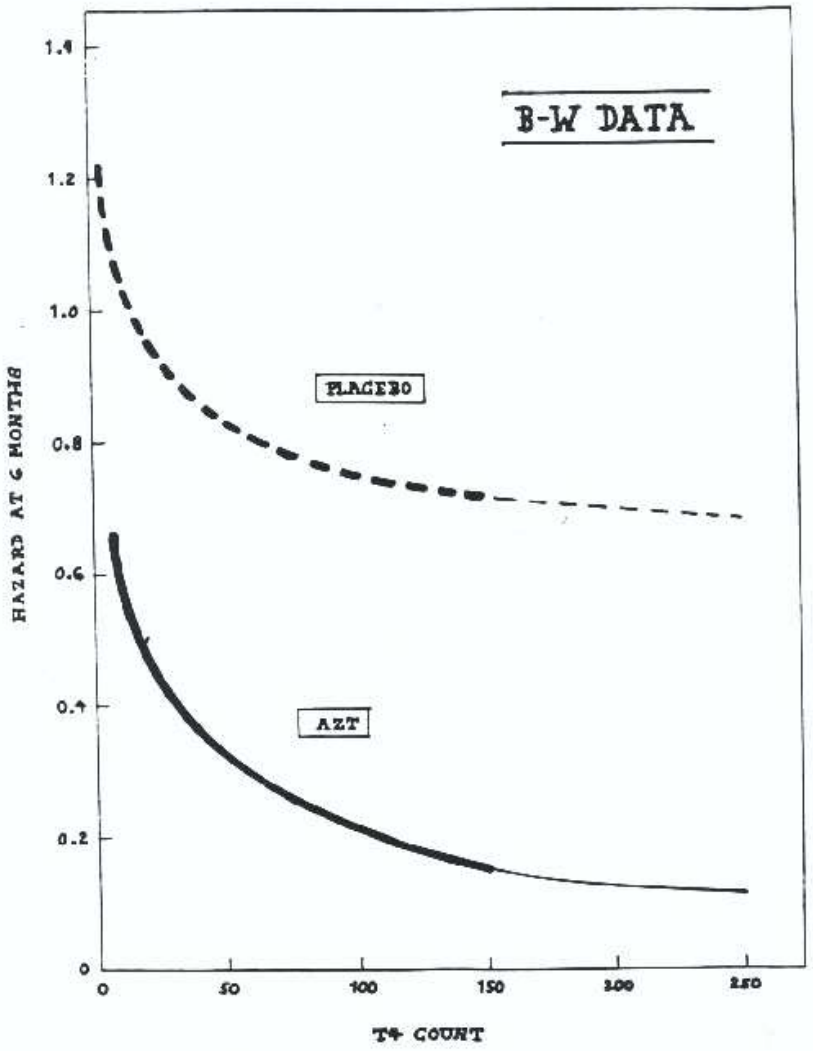
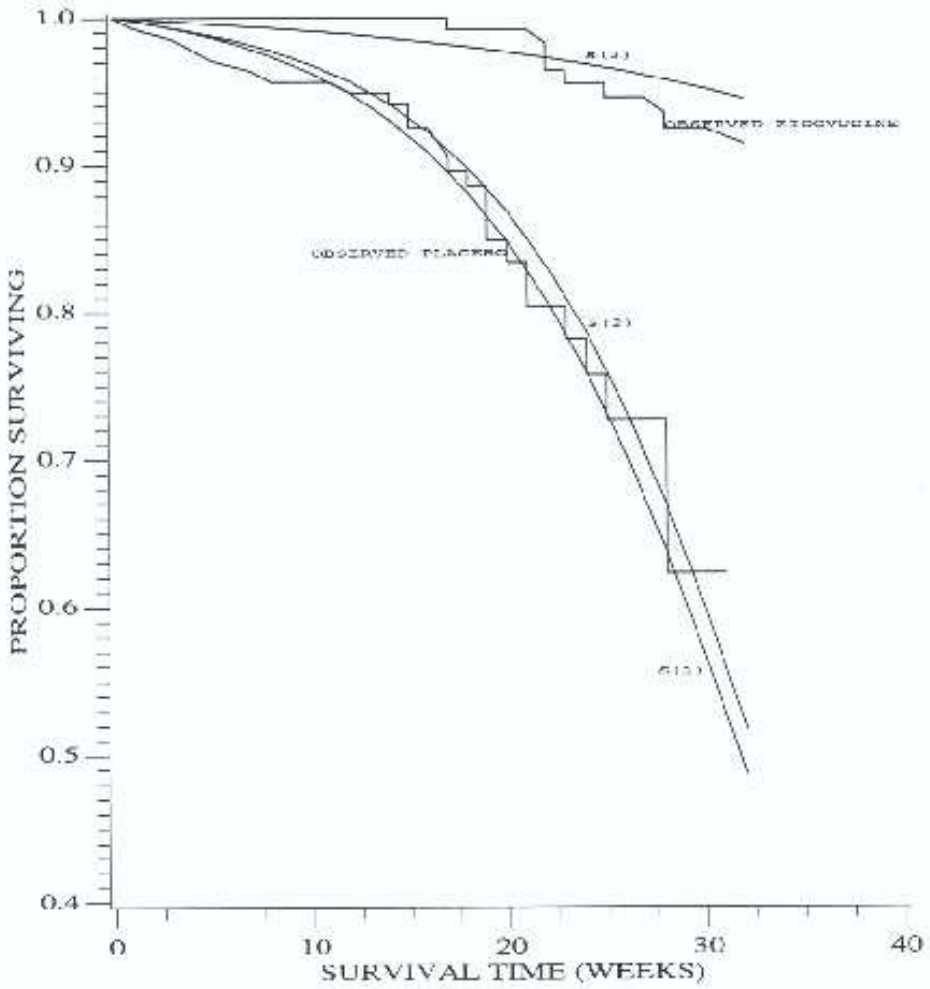

B-W DATA
PLACEBO
AZT
HAZARD AT 6 MONTHS
T4 COUNT
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250

Figure 7


1.0
0.9
0.8
0.7
0.6
0.5
0.4
PROPORTION SURVIVING
0
10
20
30
40
SURVIVAL TIME (WEEKS)

Conclusions

- treatment effects CD4 counts
- treatment effect survival
- CD4 count is a good prognostic variable
- CD4 counts can only explain a small part of the drug's clinical effect

Interim Monitoring and Group-Sequential Methods

- Comparison of two or more treatments are made throughout the course of the trial (Interim Analyses)
- Depending on the strength or weakness of the treatment comparisons, the trial may be stopped
  - A winner declared
  - no difference (futility)
- One of more treatments may be dropped and the study continued until the next interim analysis or until Final Analysis

Strategy for Decisions

- What are the criteria for deciding when treatment differences are sufficiently large or small at an interim analysis to warrant action
- Frequentist vs Bayesian
- Frequentist is concerned about operating characteristics of decision rule
  - Type I error (probability of declaring treatment difference when there is none)
  - Type II error (probability of not declaring treatment difference when there is one)

Frequentist Continued

- Because there are multiplicity of decisions to be made over time one must be worried about inflated error probabilities
- For example, if you stopped the first time the p-value was less than .05, then the chance of finding a significant difference by chance alone (even if there is none) is increased artificially when many treatment comparisons are made over time
- Frequentist stopping rules take this into account
  - O'Brien-Fleming boundaries
  - Conditional power calculations (Given the current data it is unlikely to find a significant result if the study continued, even under the alternative hypothesis)

Frequentist Continued

- Nonetheless, Group-sequential designs will use (on average) a smaller sample size to reach conclusion with the same precision as fixed sample size designs.
- The gains can be from 10%-40% decrease in average sample size
- Efficient Designs (Jennison and Turnbull)

Bayesian Methods (Peter Thall)

- A prior distribution is assumed for the parameter of interest (usually a weak prior is used)
- Let Δ denote population treatment difference

  $\Delta = 0$ corresponds to no treatment difference

  $\Delta > 0$ new treatment is effective

Bayesian Methods

- Decisions are based
  - Posterior distribution of Δ given data; e.g. one strategy may be if $P(\Delta > 0|\text{data})$ is large; i.e. if the probability that treatment is better than control is large, then stop the study
  - Predictive probability; i.e. Given the current data what is the probability of seeing a significant result at some later time? If this is large, stop the study
  - Although similar to conditional power, the probability is computed based on the best estimate of the distribution of the parameter using the available data

Predictive Probability

- With a weak prior the predictive probability is a function of current p-value and the horizon for the prediction
- Suppose we want to predict when we have K times the amount of data that are currently available with a p-value p, then the predictive probability equals

$$\Phi\left\{\Phi^{-1}(1-p)\left(1+\frac{1}{K}\right)^{1/2}-1.96\left(\frac{1}{K}\right)^{1/2}\right\}$$

- When K gets large, then the predictive probability approaches

$$1-\text{p-value}$$

Bayesian continued

- Both the posterior probability $P(\Delta > 0|\text{data})$ and the predictive probability, in the limit, are approximately equal to

$$1 - \text{p-value}$$

- Roughly this would correspond to stopping the study the first time the p-value is sufficiently small
- Does not account for multiple comparisons
- In fairness, Peter evaluates the type I and type II errors of his strategies using simulation; hence not much different from frequentist methods

Auxiliary Covariates

- Another way of gaining efficiency is by using auxiliary covariates; i.e. variables correlated with outcome
- Baseline covariates can enrich treatment comparisons due to loss of information because of randomization
- Baseline as well as post-treatment covariates can enrich treatment comparisons due to loss of information because of censoring

Randomization

- Say you are comparing a new treatment to control in a randomized study
- Even though randomization has a great deal of advantage, one disadvantage is that you don't use all the patients in a study to learn about the new treatment (half the patients are given control); i.e. you're losing half the information on the new treatment.
- However, if you have collected auxiliary covariates at baseline that are correlated with the response to the new treatment, then you can recover some of the information lost on the new treatment from those patients randomized to the control group by using the auxiliary covariates from those patients

Randomization continued

- Similarly, we can use the auxiliary covariates from those patients randomized to the new treatment to recover some of the response to placebo data that are not observed for those patients
- This is what is behind covariate adjustment and efficient methods of Davidian, Tsiatis and Leon using pretest-posttest data.

Censoring

- When time to event is used as primary endpoint, then endpoint data are censored for studies with limited follow-up
- However, if auxiliary covariates are collected, then relationships between the auxiliary covariates and endpoint data that are observed among those patients not censored can be used to recover some of the information lost from those censored using the auxiliary information from the censored patients.

Censoring

- This is what Peter Thall did in one of his examples where he partitioned his data according to levels of local control to gain information on survival for breast cancer patients
- Critique
  - parametric models assumed (exponential distribution by local control strata)
  - proportional effect of treatment assumed throughout and used as part of future prediction

Censoring

- Nonparametric and semiparametric methods also can be used to increase efficiency with auxiliary covariates
- The gains of efficiency are not that great unless you make strong assumptions or unless that auxiliary covariates are strongly correlated.

Comments on Jeff Mahon presentation

- Excellent overview summarizing the difficulties faced by TrialNet
- A thoughtful look at the relationship of sample size and desired effect size
- Argues for considering trials powered to detect larger effect sizes

Consequences of Running Many Small Trials Powered to Detect Large Effects

- Consider the following hypothetical scenario:
  - 70% of treatments have no effect
  - 25% of treatments have moderate effect
  - 5% of treatment have large effect
- Consider two strategies
  - Run larger studies that have 80% power to detect moderate effects at the .05 level of significance
  - Run smaller studies that have 80% power to detect large effects at the .05 level of significance

Consider Significant Results from Studies Powered to Detect Large Differences

- $.70 \times .05 = .0350$ of all treatments will be both statistically significant and have no effect
- $.25 \times .35 = .0875$ of all treatments will be both statistically significant and have moderate effect
- $.05 \times .80 = .0400$ of all treatments will be both statistically significant and have large effect
- Hence among the treatments found to be statistically significant

$$\frac{.0350}{.0350 + .0875 + .0400} = .22$$

will be treatments with no effect

Focus on Significant Results

Table 1: Among Significant Results, the Percentage With No Effect, Moderate Effect and Large Effect

	Small studies	Large Studies
No Effect	22%	12%
Moderate Effect	54%	70%
Large Effect	24%	18%

Powering for Large Effects

- Perhaps a better strategy would be to use group-sequential methods
- Group-sequential methods can be designed that have power to detect moderate treatment effects but will stop as early as possible when treatment effects are large

Concluding Remarks

- Different methods have been explored which can make designs more efficient; that is, assess treatment differences more quickly with fewer patients
- Group-sequential methods and the use of auxiliary variables can help gain efficiency
  - but the gains are modest; i.e. can decrease the sample size by 5-40%
- The big gains can be obtained by use of surrogate markers or looking only for large treatment effects
  - Greater risk of misleading results